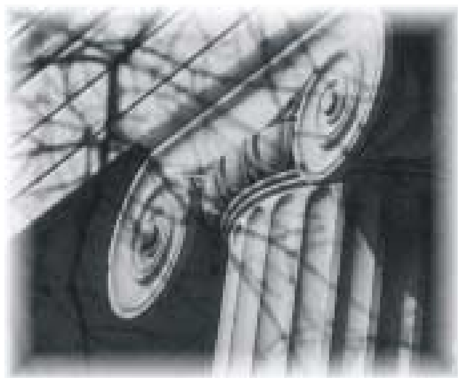
THE UNIVERSITY OF ALABAMA

ECONOMICS, FINANCE AND LEGAL STUDIES

WORKING PAPER SERIES



Working Paper No. 01-02-03

Improved Critical Values for the Enders-Granger Test

Walter Enders

Culverhouse College of Commerce and Business Administration
The University of Alabama

Improved Critical Values for the Enders-Granger Unit-Root Test*

Walter Enders
Department of Economics, Finance and Legal Studies
University of Alabama
Tuscaloosa, AL
35487

wenders@cba.ua.edu

ABSTRACT

Enders and Granger provide critical values to test the null hypothesis of a unit-root against the alternative of threshold adjustment. However, in obtaining their critical values, Enders and Granger did not use a consistent estimate of the threshold nor did they use a lag-augmented data generating process. This note remedies both of these problems. The power of the test statistics using the consistent estimate of the threshold are compared to those of Enders-Granger and of Dickey-Fuller. Surprisingly, the original Enders-Granger statistic often has the highest power. As such, the Enders-Granger statistic using a lag-augmented data generating process is calculated.

JEL Classifications: E43, C22, C50
Keywords: Asymmetric Adjustment, Monte Carlo, Nonlinear Autoregression

1. Introduction

Standard time-series models assume linearity and symmetric adjustment. Consider the simple linear relationship used as the basis for the Dickey-Fuller test:

$$\Delta y_t = \rho y_{t-1} + \varepsilon_t \tag{1}$$

where: ε_t is an identically and independently distributed (i.i.d.) disturbance term.

The standard procedure is to estimate ρ and to ascertain whether $\rho^2 < 1$ using the appropriate critical values. Although (1) can be augmented with deterministic regressors and lagged changes of $\{y_t\}$, the key point is that the adjustment process is assumed to be symmetric around $y_t = 0$. This can be problematic since Pippenger and Goering (1993), Balke and Fomby (1997), Enders and Granger (1998) and Enders and Siklos (1998) all show that traditional tests for unit-roots and cointegration have low power in the presence of asymmetric adjustment.

One way to circumvent this problem is to specify a particular non-linear adjustment mechanism and to test the null hypothesis of a unit-root against the alternative of the specific adjustment mechanism. Towards this end, Enders and Granger (1998) consider the specific threshold autoregressive (TAR) model:

$$\Delta y_t = I_t \rho_1 (y_{t-1} - \tau) + (1 - I_t)\rho_2 (y_{t-1} - \tau) + \varepsilon_t \tag{2}$$

where: I_t is the Heaviside indicator function such that:

$$I_t = \begin{cases} 1 & \text{if } y_{t-1} \geq \tau \\ 0 & \text{if } y_{t-1} < \tau \end{cases} \tag{3}$$

and: τ = the value of the threshold.

Petrucelli and Woolford (1984) show that the necessary and sufficient conditions for the stationarity of $\{y_t\}$ is: $\rho_1 < 0$, $\rho_2 < 0$ and $(1 + \rho_1)(1 + \rho_2) < 1$ for any value of τ. If these conditions are met, $y_t = \tau$ can be considered the long-run equilibrium value of $\{y_t\}$. As such, (2) and (3) can capture the behavior of time-series variables that react differently to positive versus negative discrepancies from their long-run equilibrium values.

Since adjustment is symmetric if $D_1 = D_2$, the Dickey-Fuller test is a special case of (2) and (3). Moreover, if the sequence is stationary, the least squares estimates of D_1 and D_2 have an asymptotic multivariate normal distribution.[1] However, under the null hypothesis of a unit-root, the estimates of D_1 and D_2 have non-standard distributions. Enders and Granger (1998) tabulate the appropriate critical values to test the null hypotheses $D_1 = D_1 = 0$.[2] Rejecting the null hypothesis implies an adjustment mechanism given by (2) and (3).

In (3), the Heaviside indicator depends on the *level* of $y_{t\text{-}1}$. Enders and Granger (1998) and Caner and Hansen (1998) suggest an alternative such that the threshold depends on the previous period's *change* in $y_{t\text{-}1}$. Consider setting the Heaviside indicator according to the following rule:

$$I_t \text{ ' } \begin{cases} 1 \quad \text{if }) \, y_{t\&1} \ \$ \ J \\ 0 \quad \text{if }) \, y_{t\&1} < J \end{cases} \qquad (4)$$

Models constructed using (2) and (4) are called momentum-threshold autoregressive (M-TAR) models in that the $\{y_t\}$ series exhibits more "momentum" in one direction than the other.

In general, the value of J is unknown and needs to be estimated along with the values of D_1 and D_2. Enders and Granger (1998) propose the following three-step procedure to test for a unit-root against the alternative of TAR or M-TAR adjustment.

Step 1: Demean the $\{y_t\}$ sequence by regressing y_t on a constant save the residuals as the $\{\hat{y}_t\}$ sequence. Set the indicator function I_t according to whether $\hat{y}_{t\text{-}1}$ (or $)\,\hat{y}_{t\text{-}1}$) is positive or negative. Estimate a regression equation in the form of (2) and record the F-statistic for the null hypothesis $\rho_1 = \rho_2 = 0$. Compare this sample statistic with the appropriate critical tabulated by Enders and Granger (1998). The empirical *F*-distribution using the TAR model is called the Φ_μ-statistic and that using M-TAR adjustment is called the Φ_μ^*-statistic.

Step 2: If the alternative hypothesis is accepted, it is possible to test for symmetric versus asymmetric adjustment since the joint distribution of ρ_1 and ρ_2 converge to a multivariate normal. As such, the restriction that adjustment is symmetric (i.e., the null hypothesis: $\rho_1 = \rho_2$) can be tested using the usual *F*-statistic.

Step 3: Diagnostic checking of the residuals should be undertaken to ascertain whether the estimated $\{\varepsilon_t\}$ series can reasonably be characterized by a white-noise process. If the residuals are correlated, return to Step 2 and re-estimate the model in the form:

$$\Delta \hat{y}_t = I_t \rho_1 \hat{y}_{t-1} + (1 - I_t)\rho_2 \hat{y}_{t-1} + \gamma_1 \Delta \hat{y}_{t-1} + ... + \gamma_p \Delta \hat{y}_{t-p} + \varepsilon_t \qquad (5)$$

Lag lengths can be determined by an analysis of the regression residuals and/or using a number of widely used model selection criteria such as the AIC or BIC.

2. A Consistent Test for TAR and M-TAR Adjustment

Note that by demeaning the $\{y_t\}$ sequence, the Enders-Granger procedure actually estimates (2) in the form:

$$\Delta y_t = I_t \rho_1 (y_{t-1} - \bar{y}) + (1 - I_t)\rho_2 (y_{t-1} - \bar{y}) + \varepsilon_t \qquad (6)$$

where: $\bar{y}$ is the sample mean.

However, if adjustment is asymmetric, the sample mean is a biased estimate of τ. For example, if $|\rho_1| > |\rho_2|$, autoregressive decay is slower for positive values of $(y_{t-1} - \tau)$ than for negative values. As such, if $|\rho_1| > |\rho_2|$, the sample mean will exceed the value of τ. Chan (1993) shows that it is possible to obtain a super-consistent estimate of the threshold by ordering the $\{y_t\}$ sequence such that $y^o_1 < y^o_2 < y^o_3 ... < y^o_T$. For each value of y^o_j, set $J = y^o_j$ and estimate an equation in the form of (2). The regression equation with the smallest residual sum of squares contains the consistent estimate of the threshold. In practice, the highest and lowest 15% of the $\{y^o_j\}$ values are excluded from the grid search so as to ensure an adequate number of observations on each side of the threshold.

To incorporate Chan's methodology into a Monte Carlo experiment, 45,000 random-walk series of the following form were generated:

$$y_t = y_{t-1} + ,_t, \quad t = 1, ..., T \qquad (7)$$

For sample sizes T = 50, 100 and 250, a simulated $\{ ,_t\}$ sequence was generated from a set of T normally distributed and uncorrelated pseudo-random numbers. Randomizing the initial value of y_1, the next T-1 values of $\{y_t\}$ were generated using (7). Each series was sorted in ascending order and the upper and lower 15% of the $\{y^o_j\}$ values were discarded. For each of the remaining $\{y^o_j\}$ values, we estimated an equation in the form of (2) and (3). The equation yielding the lowest residual sum of squares was deemed to be the appropriate estimate of the threshold. Using this threshold value, we obtained the F-statistic for the null hypothesis for $D_1 = D_2 = 0$. This process was repeated 45,000 times and the distribution of the resulting F-statistic--called the M_μ(c) statistic--is shown in Table 1 for various sample sizes and lag lengths. Table 2 performs the identical experiment using the momentum model. The distribution of the M^*_μ(c) statistic is reported in Table 2.[3] For example, using T = 100, Table 1 shows that the M_μ(c) statistic for the null hypothesis $D_1 = D_2 = 0$ exceeded 6.00 in approximately 5% of the 45,000 trials and Table 2 shows that the M^*_μ(c) statistic exceeded 5.77 in approximately 5% of the trials.

3. Power Tests

Since unit-root tests suffer from low power, it is of interest to compare the power of the $M_\mu(c)$ and $M^*_\mu(c)$ statistics to those of the Enders-Granger M_μ and M^*_μ statistics and to the more traditional Dickey-Fuller test. Toward this end, 100 normally distributed i.i.d. random numbers were drawn to represent the $\{ ,_t\}$ sequence. For various values of D_1 and D_2, these random numbers were used to generate the basic TAR model given by (2) and (3) for a true value of $J = 0$ (Note that the power functions are independent of the true value of J). The resulting series was tested for a unit-root using the Dickey-Fuller test, the Enders-Granger M_μ-statistic and the $M_\mu(c)$ statistic calculated above. This process was repeated 2500 times and the percentage of instances in which each test correctly rejected the null hypothesis of a unit-root is reported in Table 3 for test sizes of 10%, 5%, and 1%. The identical process was repeated using M-TAR adjustment; the results are reported in Table 4.

The overwhelming impression is that the power of the Dickey-Fuller exceeds that of the M_μ and $M_\mu(c)$ statistics at most reasonable ranges of asymmetry. For example, if the true adjustment parameters are D_1 = -0.05 and D_2 = -0.20, at the 5% significance level the M_μ and $M_\mu(c)$ statistics correctly identified the model as stationary in 26.76% and 20.20% of the trials, respectively. However, for the same sized test, the Dickey-Fuller test correctly identified the model as stationary in 31.12% of the trials. This result reinforces that of Enders and Granger (1998)--if TAR adjustment is suspected, use the Dickey-Fuller test to establish the presence (or absence) of a unit-root. The explanation lies in the fact that the TAR model entails the estimation of an additional coefficient(s) with a consequent loss of power. For plausible degrees of asymmetry, the gain in power resulting from estimating a correctly specified model does not outweigh the loss from the additional coefficient. It is interesting to note that the M_μ-statistic never has the greatest power. However, depending on the values of D_1 and D_2, it can have more power than the Dickey-Fuller test or the M_μ(c)-statistic.

The situation is quite different for the M-TAR model. Inspection of Table 4 shows:

1. If adjustment is nearly symmetric (such that D_1 . D_2), the power of the Dickey-Fuller test exceeds that of the M_μ^*(c) test. When adjustment is nearly symmetric, the assumption of asymmetric adjustment entails the needless estimation of an additional coefficient with a consequent loss of power. Increasing the degree of asymmetry increases the relative power of the M_μ^*(c) test.

2. For a reasonable range of asymmetry, the power of the M_μ^*-test exceeds that of the Dickey-Fuller test. For example, if we use a test size of 5% and set D_1 = -0.025 and D_2 = -0.20, the M_μ^*-test correctly rejects the null hypothesis of no unit-root in 58.32% of the trials. In contrast, the Dickey-Fuller is correct in only 34.28% of the instances.

3. In all instances, the power of the M_μ^*-test exceeds that of the M_μ^*(c) test statistic. For example, at the 5% significance level, if D_1 = -0.025 and D_2 = -0.20, the M_μ^*-test correctly rejects the null hypothesis of no unit-root in 58.32% of the trials whereas the M_μ^*(c)-test is correct in only 52.32% of the instances. It is somewhat surprising that the M_μ^*-test has more power than the model using the consistent estimator of the threshold. The reason is that the bias in using the sample mean to estimate the threshold small with M-TAR adjustment. Even in an extreme case such that D_1 = -0.5 and D_2 = -0.50, the estimated mean over the 2500 trials is 1.61 if there is TAR adjustment but only 0.38 with M-TAR adjustment.

Given the power of the M_μ^*-statistic, it seems reasonable for the researcher to use this test if M-TAR adjustment is suspected. However, the distribution of this statistic depends on the number of lagged changes of {) y_t} used estimating equation. Enders and Granger (1998) do not report results a dynamic adjustment equation in the form of (4). For convenience, Table 5 contains the distribution of the Enders-Granger M_μ^*-statistic when 0, 1 and 4 lagged changes of {) y_t} are included in the dynamic adjustment process.[4]

4. Conclusions and Recommendations

The standard tests for a unit-root implicitly assume that adjustment to the long-run equilibrium relationship is symmetric. This note extends the Enders-Granger (1998) unit-root tests that allow for threshold autoregressive (TAR) or momentum-TAR (M-TAR) adjustment under the alternative hypothesis. Critical values are provided for the case in which a consistent estimator of the threshold is used in the test procedure. However, the power of the tests for TAR adjustment with and without consistent estimates of the threshold are poor compared to that of the Dickey-Fuller test. The recommendation is to use the Dickey-Fuller test if TAR adjustment is suspected. However, for a plausible range of the adjustment parameters, the power of the test for M-TAR adjustment is best conducted using the sample mean as the estimate of the threshold. Since the appropriate test statistics depend on lag length, but are not found in Enders and Granger (1998), the appropriate critical values are reported here.

References

Balke, N.S. and T. Fomby (1997). "Threshold Cointegration." *International Economic Review* 38, 627 - 43.

Caner, Mehmet and Bruce Hansen (1998). "Threshold Autoregression with a Near Unit Root." University of Wisconsin Working Paper: *mimeo*.

Chan, K. S. (1993). "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model." *The Annals of Statistics* 21, 520 - 33.

Enders, Walter and C.W.J. Granger (1998). "Unit-Root Tests and Asymmetric Adjustment With an Example Using the Term Structure of Interest Rates." *Journal of Business and Economic Statistics* 16, 304 - 11.

_____ and Pierre Siklos (1998). "Cointegration and Threshold Adjustment." Department of Economics Working Paper: Iowa State University. *mimeo*.

Petrucelli, J. and S. Woolford (1984). A Threshold AR(1) Model. *Journal of Applied Probability* 21, 270 - 86.

Pippenger, M. K. and G. E. Goering (1993), "A Note on the Empirical Power of Unit Root Tests Under Threshold Processes." *Oxford Bulletin of Economics and Statistics* 55, 473 - 81.

Tong, Howell (1990). *Non-Linear Time-Series: A Dynamical Approach*. (Oxford University Press: Oxford).

_________ (1983). *Threshold Models in Non-Linear Time Series Analysis*. (Springer-Verlag: New York).

Table 1: The Distribution of the $M_\mu(c)$-Statistic

	No Lagged Changes				One Lagged Change				Four Lagged Changes			
	90%	95%	97.5	99%	90%	95%	97.5	99%	90%	95%	97.5	99%
50	5.15	6.19	7.25	8.64	5.55	6.62	7.66	9.10	5.49	6.55	7.59	9.00
100	5.08	6.06	6.93	8.19	5.39	6.34	7.30	8.54	5.38	6.32	7.29	8.56
250	5.11	6.03	6.88	8.04	5.26	6.12	6.99	8.14	5.36	6.29	7.15	8.35

Table 2: The Distribution of the $M^*_\mu(c)$-Statistic

	No Lagged Changes				One Lagged Change				Four Lagged Changes			
	90%	95%	97.5	99%	90%	95%	97.5	99%	90%	95%	97.5	99%
50	5.02	6.05	7.09	8.59	4.98	6.07	7.15	8.56	4.93	5.96	7.01	8.48
100	4.81	5.77	6.73	7.99	4.77	5.71	6.56	7.90	4.74	5.70	6.67	7.97
250	4.70	5.64	6.51	7.64	4.64	5.54	6.40	7.56	4.64	5.54	6.39	7.61

Table 3 : Power Comparisons Under TAR Adjustment

		The M_μ-Statistic			**The Dickey-Fuller Test**			**The $M_\mu(c)$-Statistic**		
D_1	D_2	10%	5%	1%	10%	5%	1%	10%	5%	1%
-0.025	-0.10	18.44	10.64	2.44	22.80*	12.68*	2.76*	15.64	8.16	2.12
-0.025	-0.20	26.84	16.12	4.44	30.56*	18.16*	4.92*	24.16	13.84	3.40
-0.025	-0.30	34.00	21.88	8.08	36.88*	23.76*	8.28	34.76	22.48	7.16
-0.025	-0.50	40.72	28.88	11.84	42.32	28.40	11.24	47.76*	35.52*	15.76*
-0.05	-0.05	19.24	10.36	2.48	22.84*	12.84*	2.64*	14.96	7.88	1.76
-0.05	-0.10	30.88	17.24	4.32	35.76*	20.56*	4.92*	24.12	12.64	2.68
-0.05	-0.20	43.48	26.76	8.72	47.76*	31.12*	9.24*	34.96	20.20	5.72
-0/05	-0.30	48.96	32.52	12.00	52.92*	35.64*	12.92*	45.16	31.16	9.92
-0.05	-0.50	61.76	45.16	19.48	62.76	46.56	19.72	63.04*	48.72*	22.92*
-0.10	-0.20	67.40	47.36	16.72	73.12*	53.24*	18.52*	54.44	35.32	10.72
-0.10	-0.30	79.28	63.04	28.60	83.04*	66.84*	29.64*	70.04	53.04	20.44
-0.10	-0.50	89.72	78.92	46.92	90.92*	80.04*	46.80*	87.92	75.60	45.04
-0.10	-0.75	93.28	85.60	58.28	93.52	84.68	56.40	94.32*	88.48*	64.96*

Note: A (*) indicates the test with the greatest power for each test size.

Table 4 : Power Comparisons Under M-TAR Adjustment

		The M^*_μ-Statistic			The Dickey-Fuller Test			The $M^*_\mu(c)$-Statistic		
D_1	D_2	10%	5%	1%	10%	5%	1%	10%	5%	1%
-0.025	-0.10	31.00*	18.76*	4.60*	26.44	14.40	2.84	26.12	14.88	3.76
-0.025	-0.20	73.96*	58.32*	23.68*	55.56	34.28	9.92	67.88	52.32	21.36
-0.025	-0.30	96.96*	91.44*	65.20*	82.00	63.24	26.08	95.04	88.80	60.44
-0.025	-0.50	100.00*	100.00*	98.72*	97.88	93.16	67.52	100.00	99.96	98.00
-0.05	-0.05	19.60	10.48	2.36	21.32*	11.12*	2.60*	14.16	7.60	1.44
-0.05	-0.10	35.92	21.88*	5.92*	36.60*	21.24	4.96	27.64	16.08	4.28
-0.05	-0.20	75.56*	57.84*	22.28*	65.84	46.36	13.36	67.40	48.40	18.88
-0.05	-0.30	95.68*	89.20*	60.20*	86.44	70.04	31.68	93.56	84.48	55.20
-0.05	-0.50	100.00*	99.95*	98.36*	99.20	96.78	78.04	100.00*	99.84	97.32
-0.10	-0.10	44.36	25.88	7.44	51.60*	31.80*	8.68*	32.84	18.96	4.60
-0.10	-0.20	77.60	58.84	23.52*	80.04*	60.64*	23.08	67.32	48.68	17.44
-0.10	-0.30	97.16*	89.52*	58.24*	95.44	84.56	46.20	93.04	82.64	49.00
-0.10	-0.50	100.00*	99.92*	98.56*	99.88	99.24	88.76	100.00	99.96	97.80

Note: A (*) indicates the test with the greatest power for each test size.

Table 5: The Distribution of the M^*_μ-Statistic

	No Lagged Changes				One Lagged Change				Four Lagged Changes			
	90%	95%	97.5	99%	90%	95%	97.5	99%	90%	95%	97.5	99%
50	4.21	5.19	6.15	7.55	4.12	5.11	6.05	7.25	3.82	4.73	5.65	6.84
100	4.11	5.04	5.96	7.10	4.08	4.97	5.87	7.06	3.81	4.72	5.63	6.83
250	4.08	4.97	5.83	6.91	4.05	4.93	5.78	6.83	3.69	4.71	5.63	6.78

Endnotes

1. Tong (1983) contains the proof that the least squares estimates of D_1 and D_2 have an asymptotic multivariate normal distribution. This result easily generalizes to higher-order autoregressive processes. Tong (1990) also develops many of the properties of the TAR model.

2. When the estimates of D_1 and D_2 are both negative, it is tempting to use the distributions of the estimated D_i to test for stationarity. However, Enders and Granger (1998) show that the distribution for the least negative value of D_i--called the *t-max* statistic--and the most negative value of D_i--called the *t-min* statistic have little power. For this reason, the *t-max* and *t-min* statistics are not considered here.

3. Note that the various M_μ-statistics should be used only in those cases where the point estimates for D_1 and D_2 both imply convergence.

4. For comparability across the various lag lengths, the M_μ^*-statistic using no lagged changes was recalculated. Nevertheless, the values reported in the table are very similar to those reported in Enders and Granger (1998).